PNAS
www.pnas.org

Supplementary Information for

Ancient DNA from Guam and the Peopling of the Pacific

Irina Pugach, Alexander Hübner, Hsiao-chun Hung, Matthias Meyer, Mike T. Carson, Mark Stoneking

Address correspondence to Mike T. Carson or Mark Stoneking
Email: mtcarson@triton.uog.edu or stoneking@eva.mpg.de

This PDF file includes:

Figures S1 to S14

SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S8

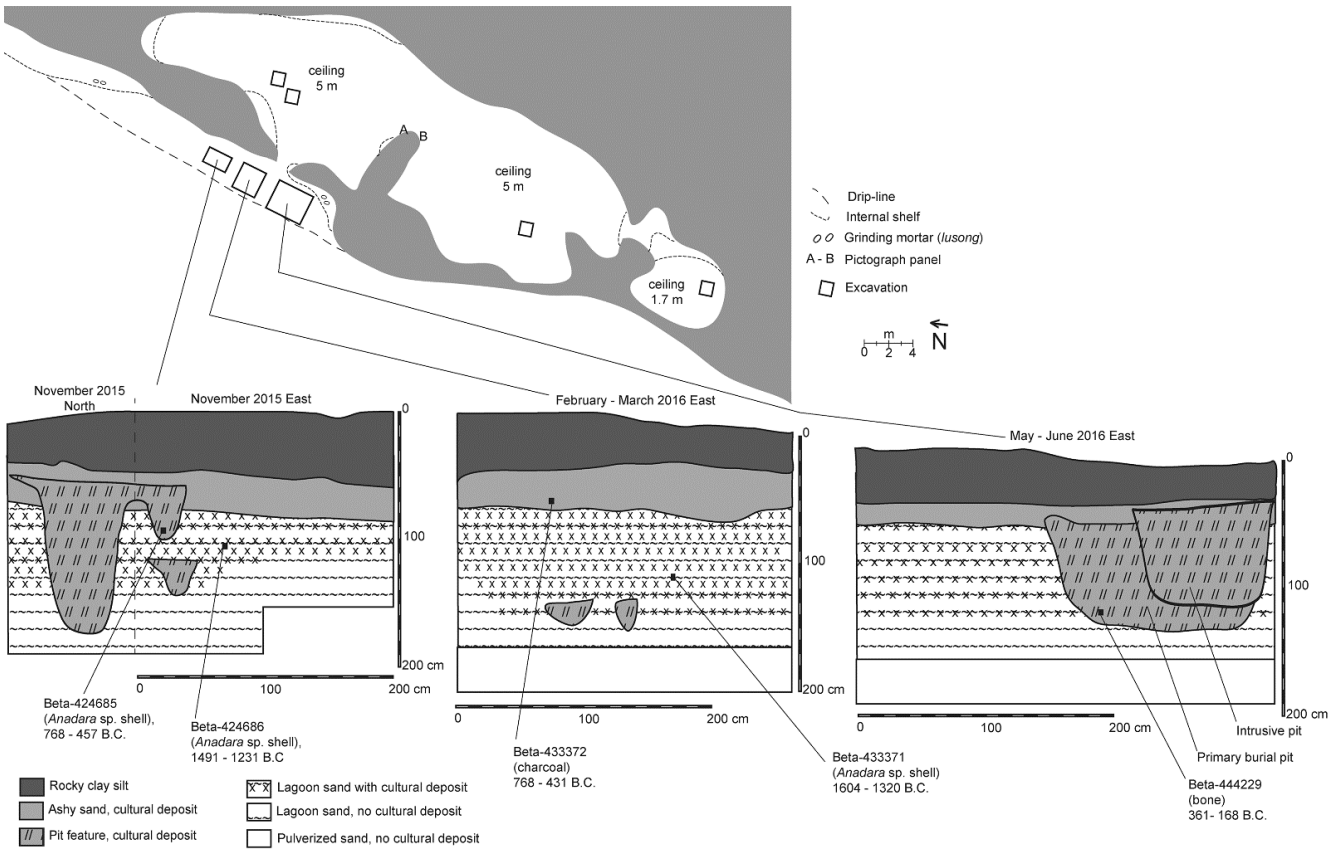


Figure S1. Plan of the Ritidian Site. Diagram modified, with permission, from (1).

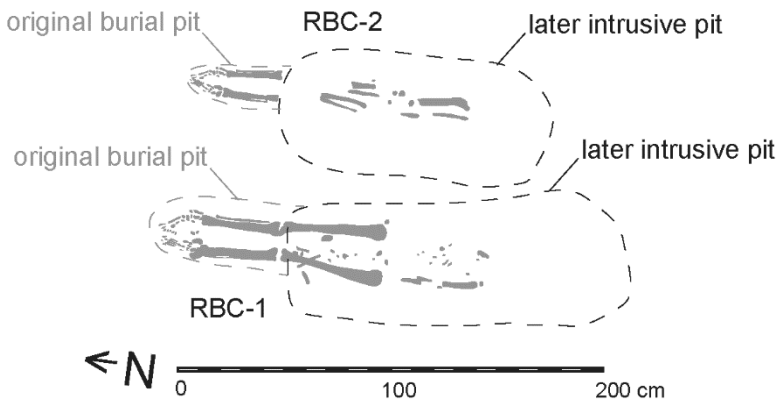


Figure S2. Diagram of the RBC-1 and RBC-2 skeletal remains and orientation of the burial pits.

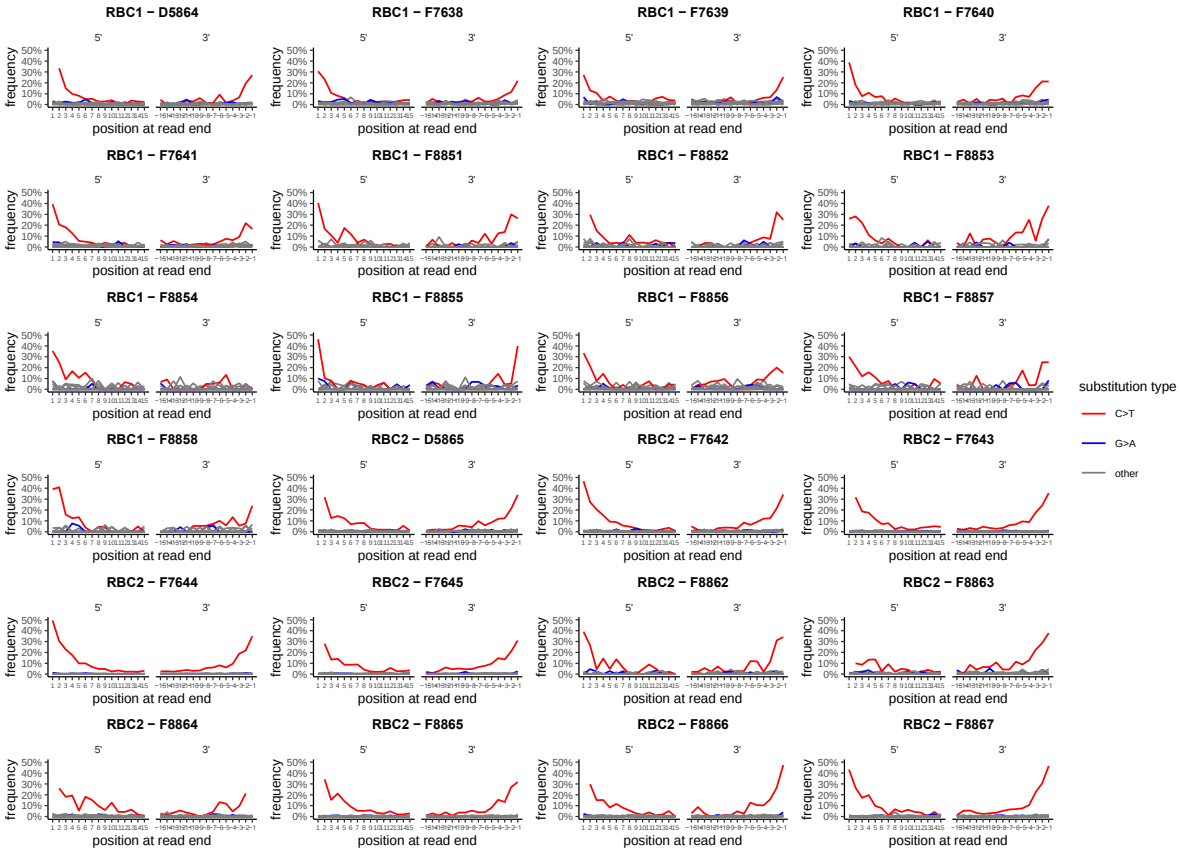


Figure S3. Substitution frequency at the terminal 15 bases of both the 5' and 3' ends of DNA fragments obtained from each library of the ancient Guam samples. The substitution frequencies were determined based on the shotgun sequencing data for each library. The substitutions C>T and G>A, which are indicative for the presence of ancient DNA damage, are highlighted in red and blue, respectively, while all other substitutions are plotted in grey. High stochasticity is caused by the low number of reads for some of the libraries (SI Appendix, Dataset S1).

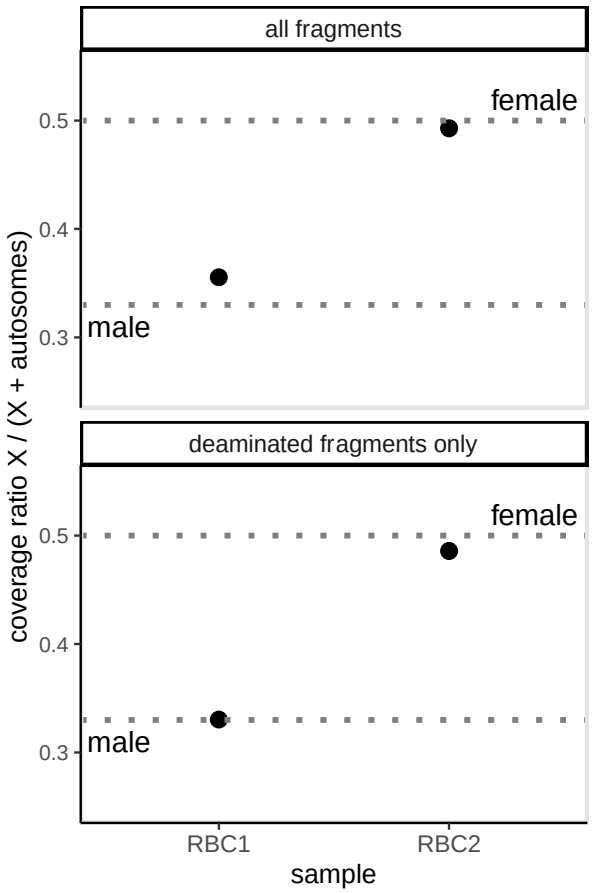


Figure S4. Ratio of X chromosome to X chromosome + autosomal reads for the ancient Guam samples, for both all fragments (top) and deaminated fragments only (bottom). The coverage across the X chromosome and the average coverage for all autosomes was determined from the shotgun sequencing. The dashed horizontal lines indicate the expected values for genetically female (0.5) vs. male (0.33) individuals. The results indicate that RBC1 is male while RBC2 is female.

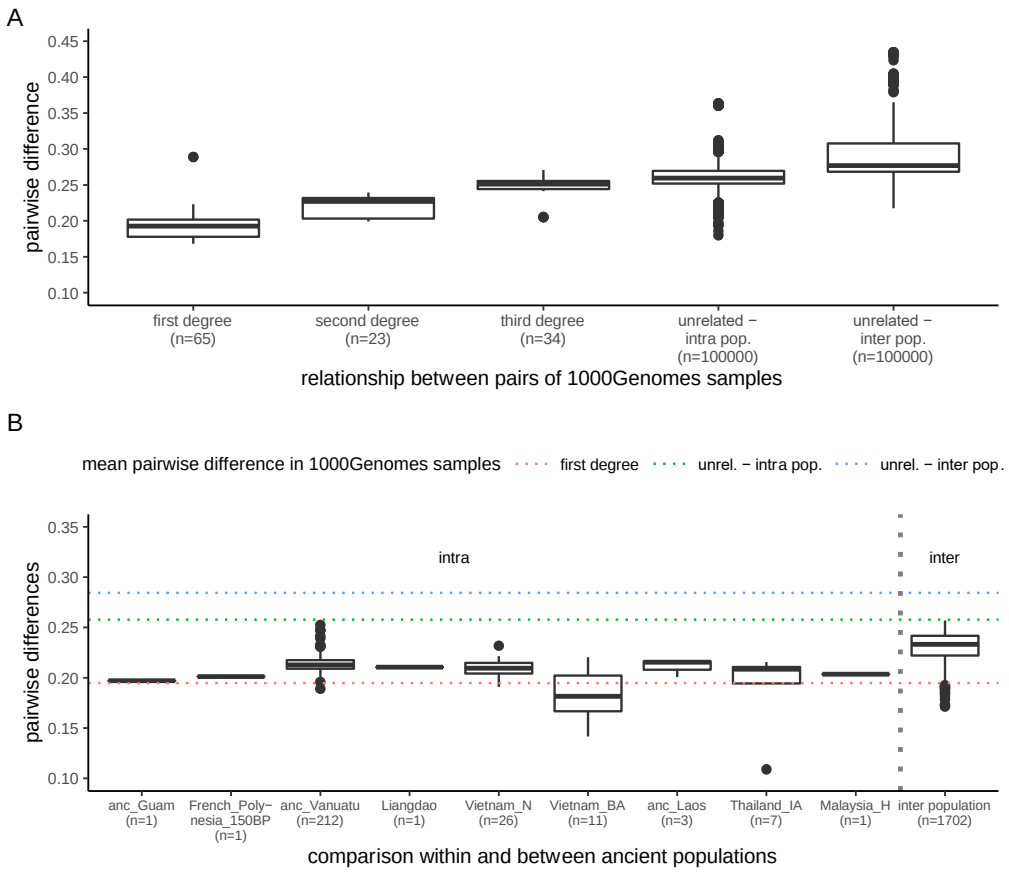
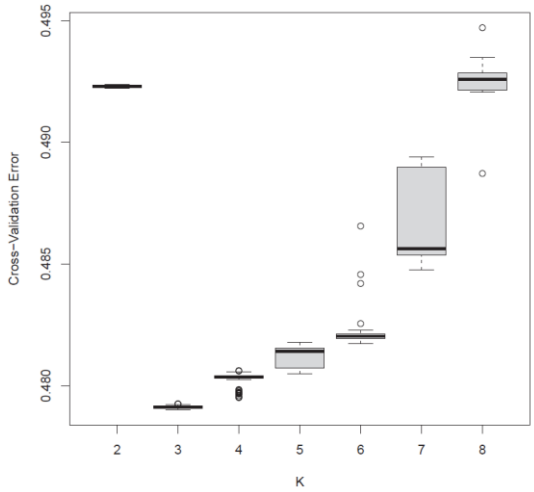


Figure S5. Comparison of the fraction of pairwise differences among present-day and ancient populations. **A** Fraction of pairwise differences among 2,535 present-day individuals from the 1000 Genomes Project data set (2) with known pairwise relatedness status. Numbers in parentheses are the number of pairwise comparisons; for unrelated individuals, only 100,000 randomly-selected pairwise comparisons within and between populations are shown to avoid a sampling depth bias. **B** Fraction of pairwise differences within and between nine ancient populations. The horizontal lines indicate the average fraction of pairwise differences that was observed in the present-day individuals from panel A for first-degree relatives and for unrelated individuals from either the same or different populations.

A.



B.

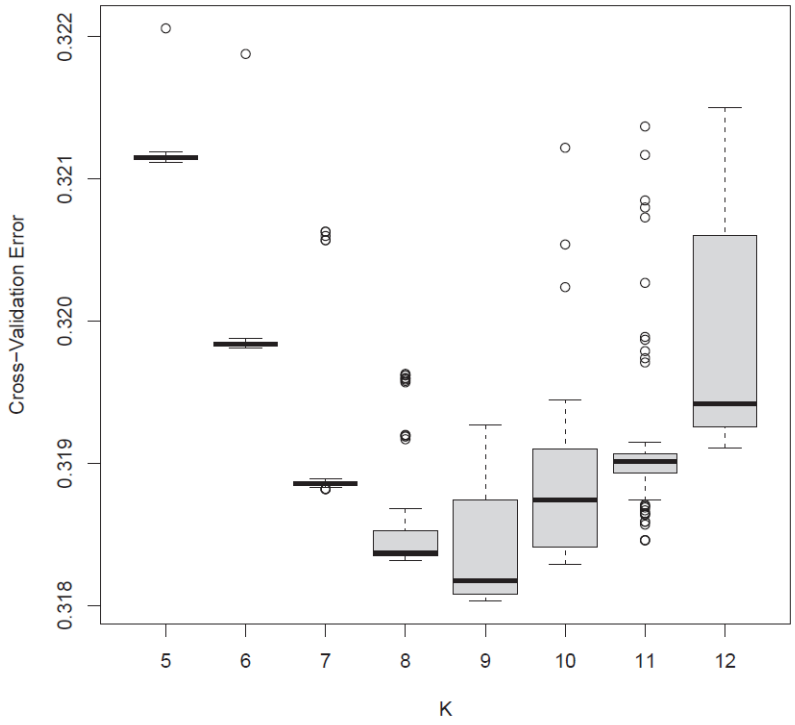


Figure S6. Plot of the cross-validation error for each value of K analyzed for ADMIXTURE analyses of: **A** the Affymetrix 6.0 + SGDP data - the results indicate that K=3 is associated with the lowest error; **B** the Human Origins array data - the results indicate that K=9 is associated with the lowest error.

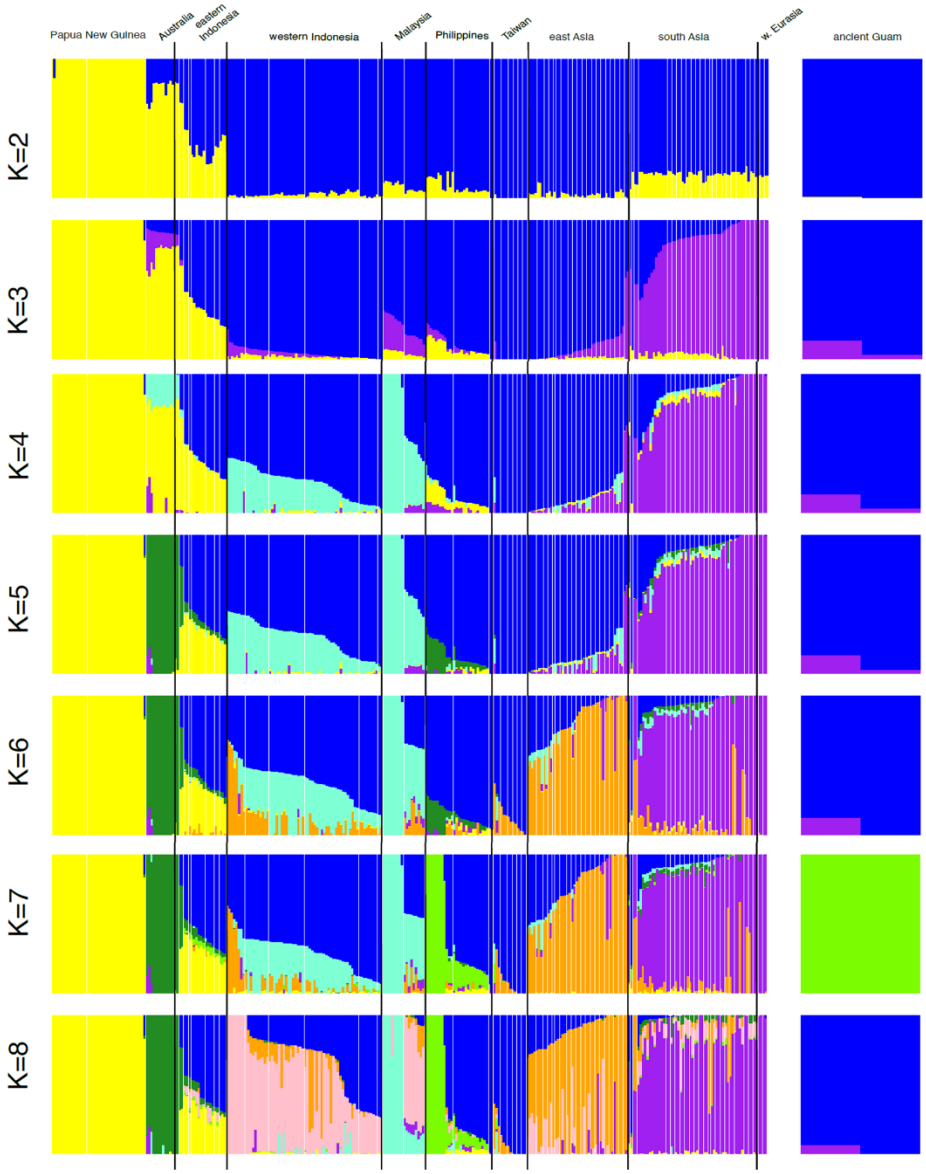


Figure S7. Plots of the results for K=2 to K=8 of the ADMIXTURE analysis of the Affymetrix 6.0+SGDP data.

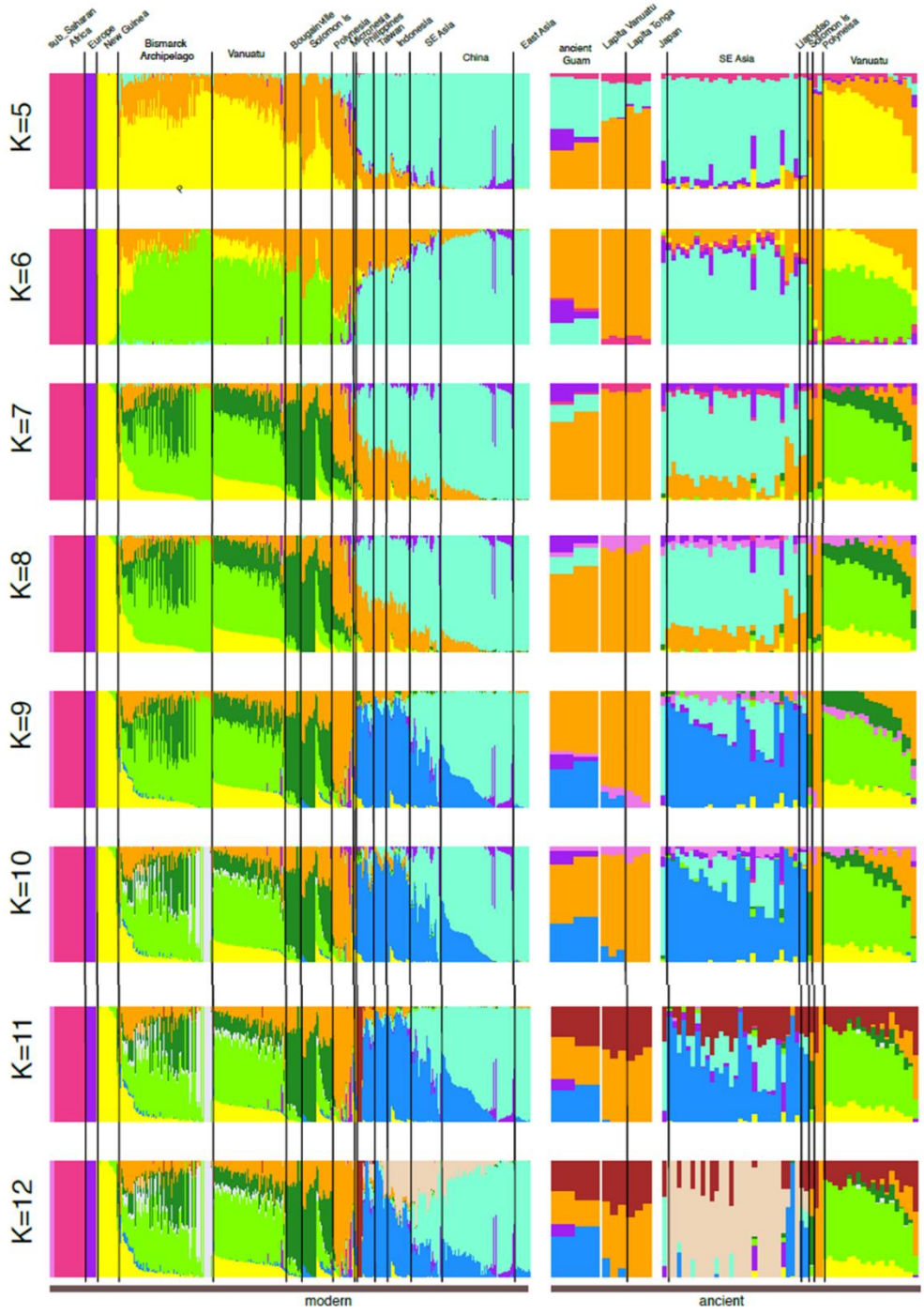


Figure S8. Plots of the results for K=5 to K=12 of the ADMIXTURE analysis of the Human Origins array data for modern and ancient samples.

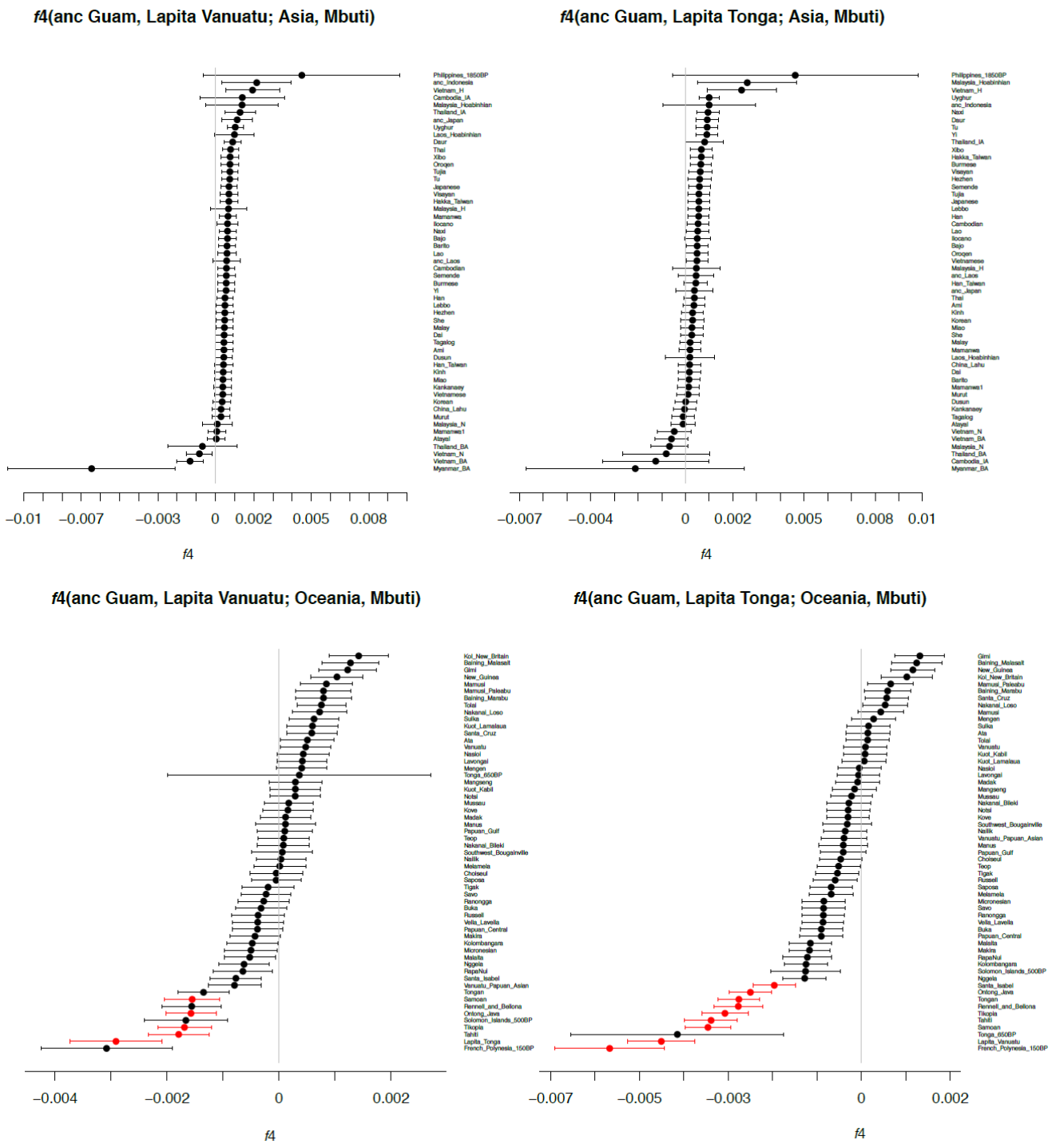


Figure S9. *f*4 statistics of the form (ancient Guam, Lapita Vanuatu/Tonga; Asia/Oceania, Mbuti), for all modern and ancient Asian and Oceanian samples. Dots indicate values of the *f*4 statistic, bars indicate 1 SE. Red dots and bars indicate values that are significantly different from zero.

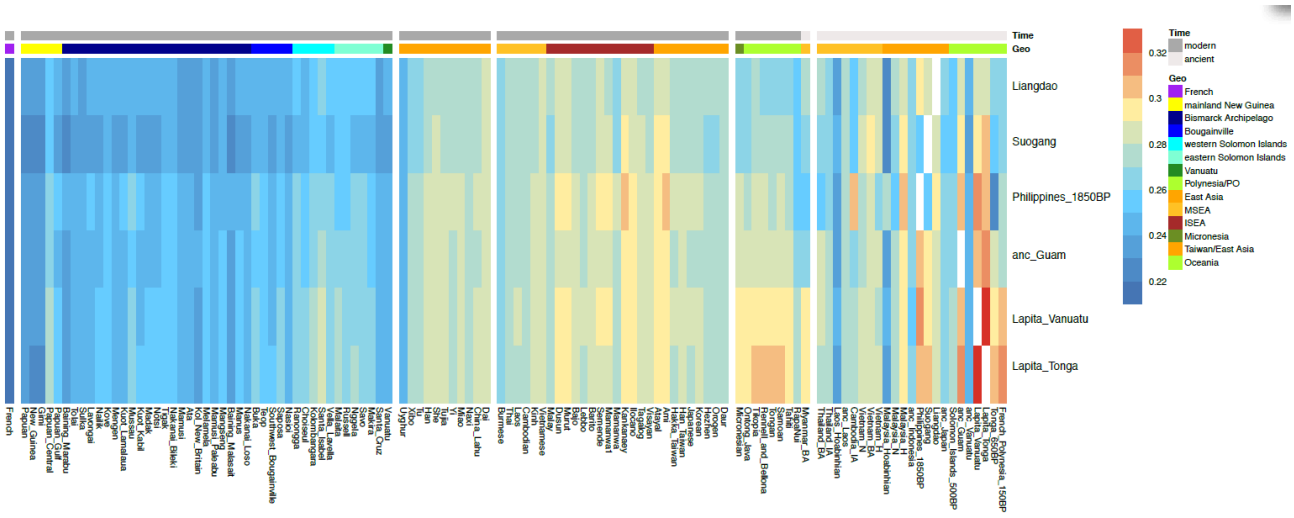


Figure S10. Heatplot of outgroup-*f*3 statistics, comparing shared drift between the early Lapita samples and other modern and ancient samples from Asia and Oceania, vs. that shared between the ancient Guam samples and the other samples. The bars above the plot indicate whether samples are modern or ancient, and their geographic origin, color-coded according to the key at right. Larger values of the outgroup-*f*3 statistic indicate more shared drift, i.e. a closer genetic relationship.

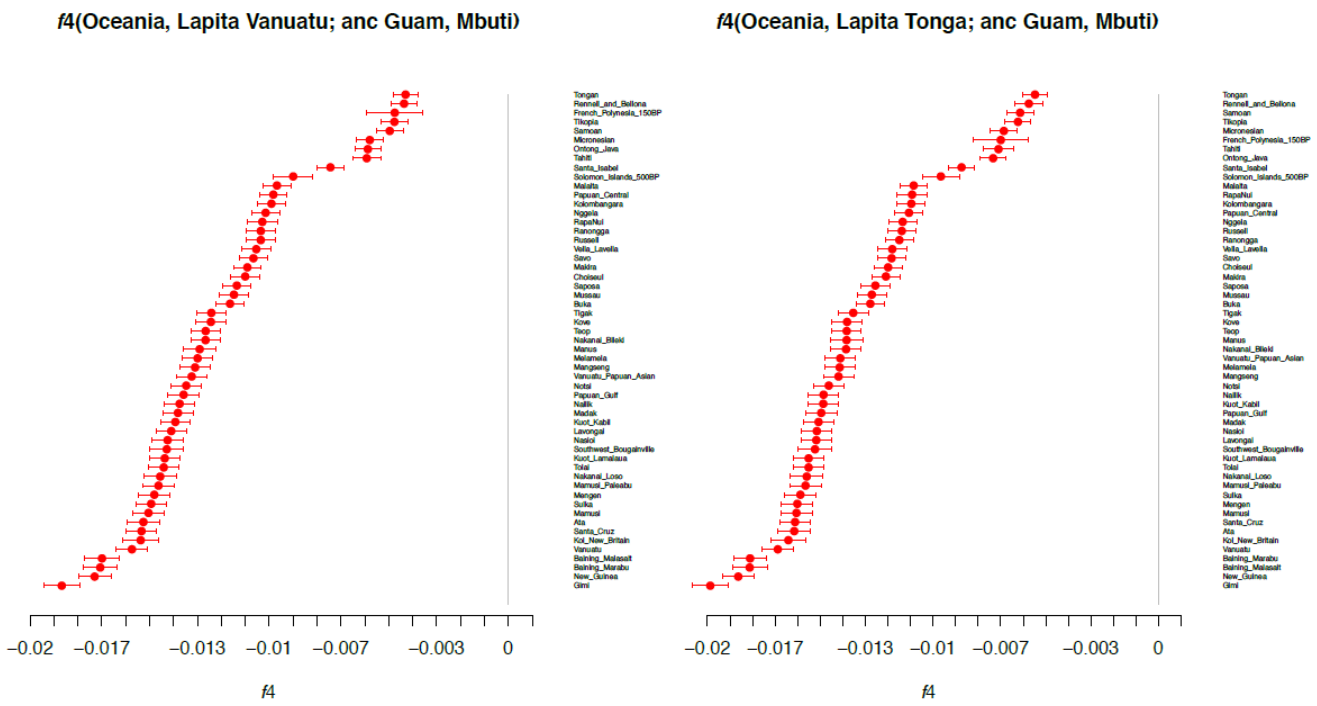


Figure S11. *f*4 statistics of the form *f*4(test, Lapita Vanuatu/Tonga; ancient Guam, Mbuti), where "test" is all modern and ancient Oceanian samples; significant results are in red, so all values are significantly different from zero. Left, Lapita Vanuatu; right, Lapita Tonga. The ancient Guam samples share excess ancestry with the early Lapita samples when compared to any other Oceanian population.

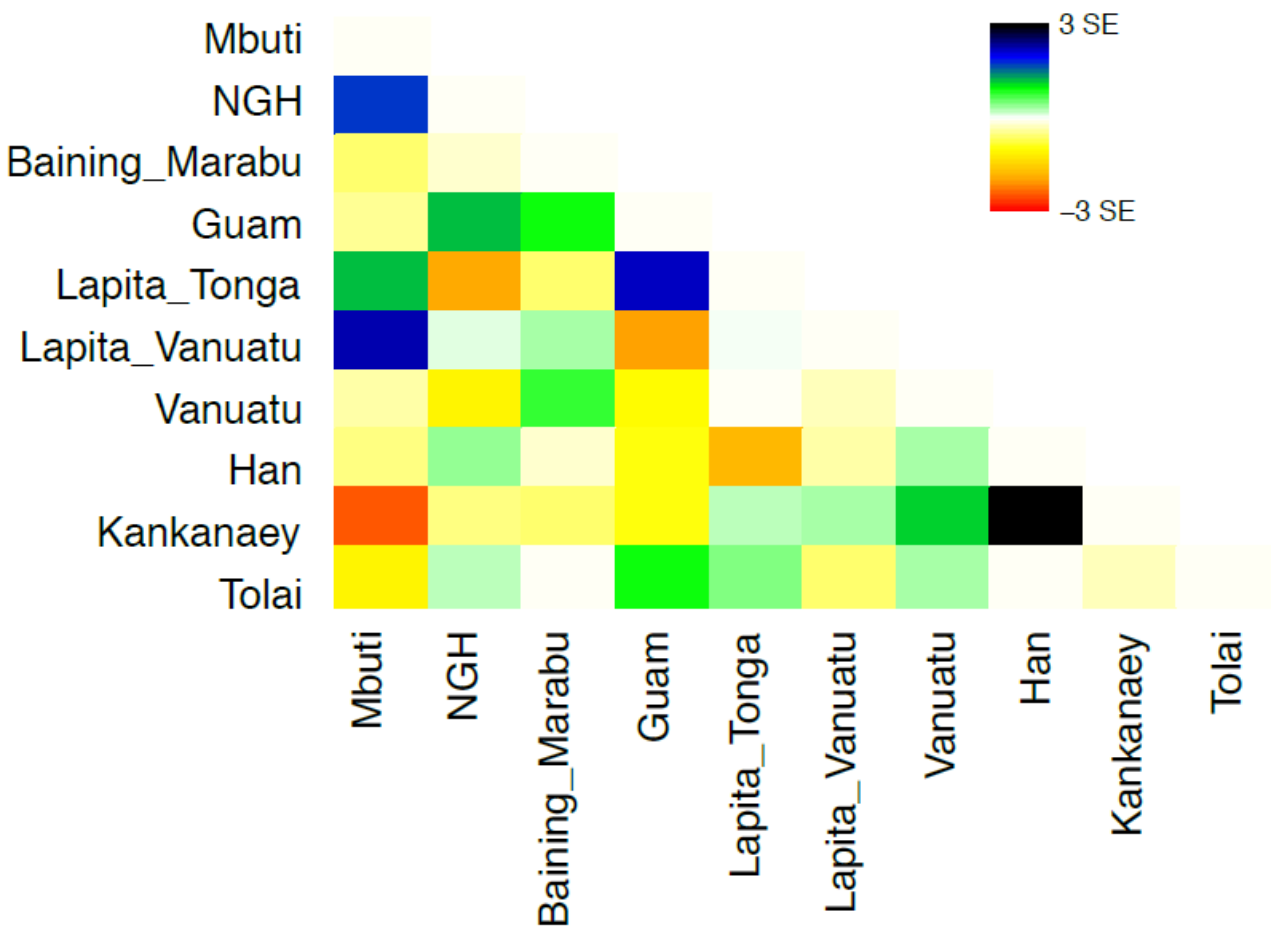


Figure S12. Residuals for the maximum-likelihood tree shown in Figure 5A, with two migration edges.

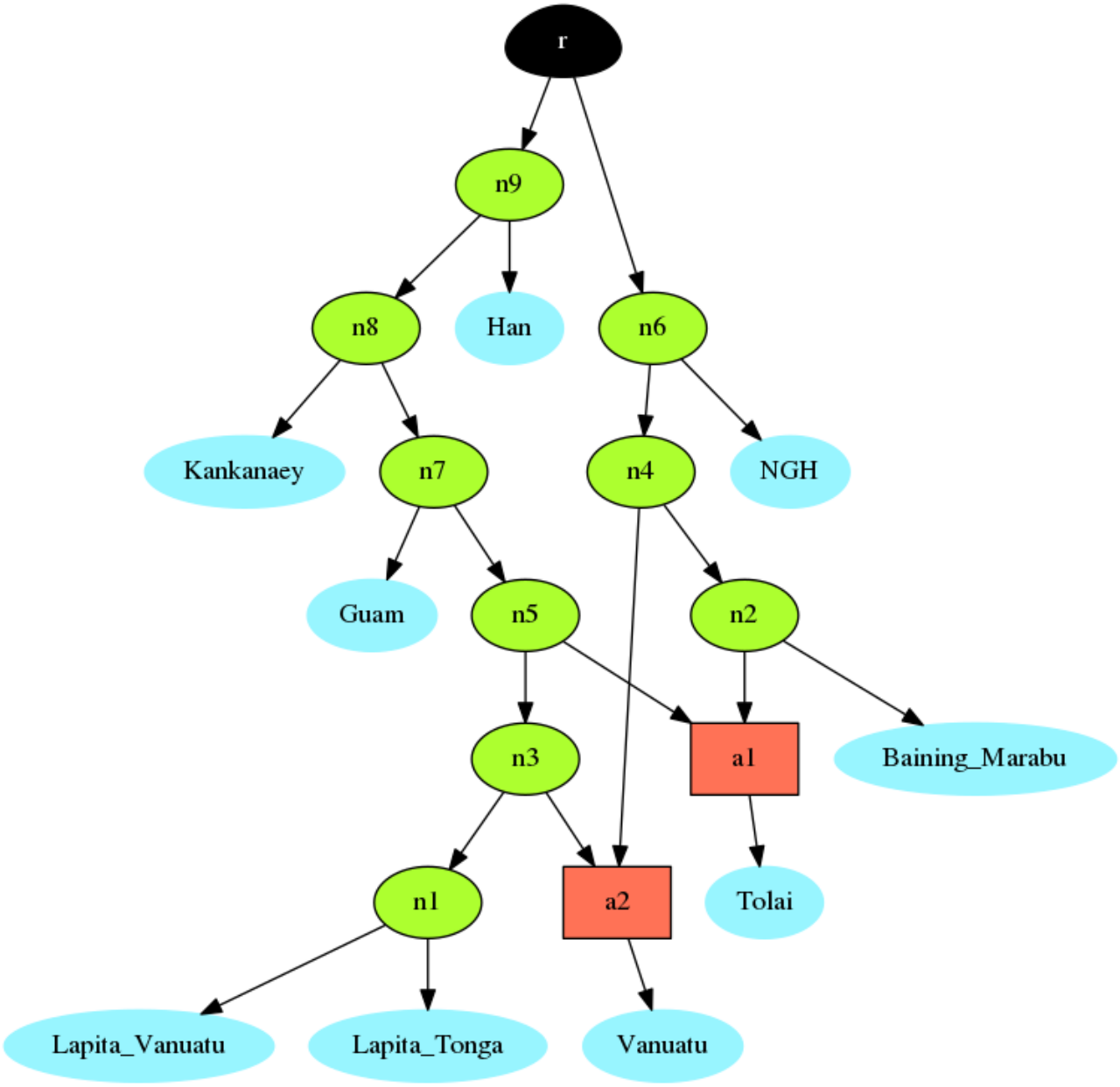
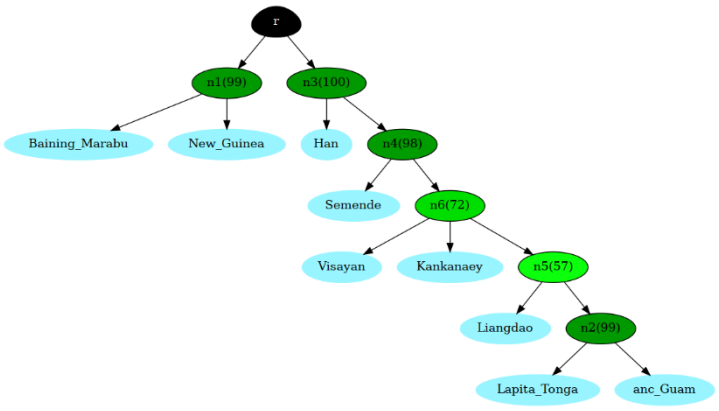


Figure S13. Topology of the graph obtained with AdmixtureBayes (3) that has the highest associated posterior probability (17.6%, vs. 7.2% for the second-best graph). Population labels are in blue, green nodes are divergence events, red nodes are admixture events, and the black r indicates the root.

A



B

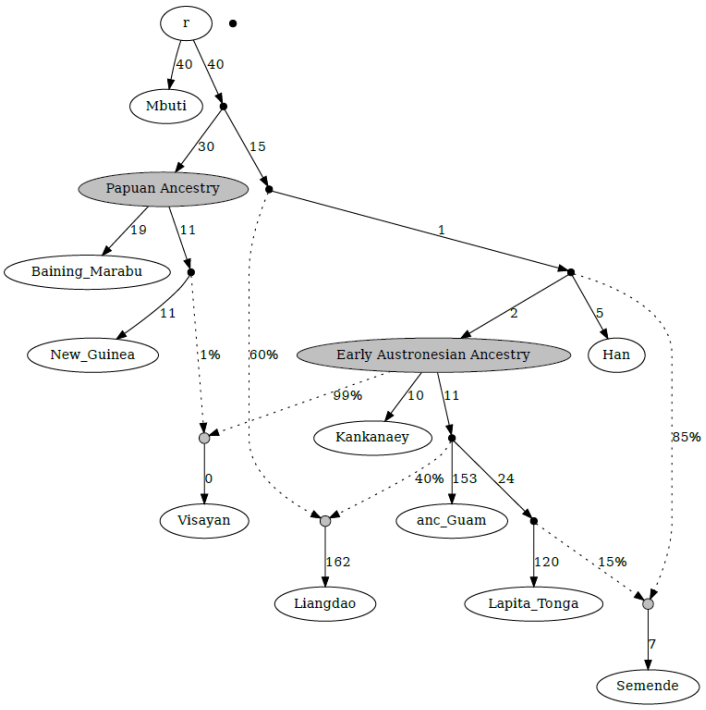


Fig. S14. Admixture graph results when including the ancient sample from Liangdao, previously shown to be related to aboriginal Taiwanese (4). **A**, consensus graph from AdmixtureBayes with nodes present in at least 50% of the topology sets depicted. **B**, best-fitting graph obtained with qpGraph for the same samples as in the AdmixtureBayes analysis, with a worst-fitting Z score of -2.585.

SI References

1. M. T. Carson, Cultural spaces inside and outside caves: a study in Guam, western Micronesia. *Antiquity* **91**, 421-441 (2017).
2. 1000 Genomes Consortium *et al.*, A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
3. S. V. Nielsen, *Inferring Gene Flow Between Populations with Statistical Methods.* (Aarhus Universitet, Aarhus, Denmark, 2018).
4. M. A. Yang *et al.*, Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* **369**, 282-288 (2020).
5. H. Weissensteiner *et al.*, HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58-W63 (2016).
6. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16** (2015).
7. K. Prufer, snpAD: an ancient DNA genotype caller. *Bioinformatics* **34**, 4165-4171 (2018).
8. G. Poznik, Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. BioRxiv: https://doi.org/10.1101/088716 (19 November 2016).
9. M. Meyer *et al.*, Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504-507 (2016).
10. S. Mallick *et al.*, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201-206 (2016).
11. D. Reich *et al.*, Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516-528 (2011).
12. A. Wollstein *et al.*, Demographic history of Oceania inferred from genome-wide data. *Curr. Biol.* **20**, 1983-1992 (2010).
13. P. F. Qin, M. Stoneking, Denisovan ancestry in East Eurasian and Native American populations. *Mol. Biol. Evol.* **32**, 2665-2674 (2015).
14. P. Skoglund *et al.*, Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510-513 (2016).
15. I. Lazaridis *et al.*, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413 (2014).
16. N. Patterson *et al.*, Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).
17. M. Lipson *et al.*, Population turnover in Remote Oceania shortly after initial settlement. *Curr. Biol.* **28**, 1157-1165.e7 (2018).
18. M. Lipson *et al.*, Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **361**, 92-95 (2018).
19. H. McColl *et al.*, The prehistoric peopling of Southeast Asia. *Science* **361**, 88-91 (2018).
20. C. Posth *et al.*, Language continuity despite population replacement in Remote Oceania. *Nat Ecol Evol* **2**, 731-740 (2018).